We would like to thank Referee #1 for their detailed and helpful review of our manuscript. We especially appreciate that the referee has volunteered additional time to read the accompanying model description paper that is published in another journal, thus further enhancing the quality of the review.

We respond to the referee's points below, whereby the referee's comments are indented.

> However, the impact of bioturbation on marine sediment records has been known for decades, and similarly, bioturbation models are around for a long time, as the study acknowledges. The work of Lougheed et al. may therefore seem marginal, but I consider the implementation of single-specimen simulations and the direct link to 14C calibration tools important.

We think this description by the referee sums up the manuscript well. Bioturbation is a well-known, but nevertheless often disregarded and misunderstood phenomenon. Its specific effect upon (calibrated) ^{14}C chronologies in deep-sea sediment cores has not yet not been systematically quantified, probably due to 20th century limitations in computing power.

> Although I have read the accompanying paper in Geoscientific Model Developments, I haven't had the time to run SEAMUS with study-independent scenarios. I assume that SEAMUS was intensively tested and run by reviewers who have assessed the accompanying paper.

Yes, the model description manuscript has now been positively assessed by two referees. Of course, no model can be conclusively guaranteed as bug-free, but we note that the output of the single-foram enabled SEAMUS model was tested in the model description paper against a well-established 'traditional' bioturbation model that considers only the mean, downcore signal (TURBO2), and the downcore mean signal output of both was found to be in virtually perfect agreement.

> Main criticism: Most of the paper deals with best-case scenarios, in which sedimentation rate, fragmentation and bioturbation depth are held constant. This rarely applies to marine sediment cores, as acknowledged. SEAMUS offers the opportunity to test the impact of bioturbation on chronological models under a variety of scenarios with transient changes of input variables, which would more closely represent marine sediment cores.

In this manuscript we chiefly seek to concentrate upon an effect that has, as far as we know, not previously been considered nor mentioned in the literature: namely age-depth artefacts that can arise as a result of the mischaracterisation of pooled foraminifera-based ^{14}C determinations as a normal distribution, coupled to the subsequent amplification of this mischaracterisation during the calibration process. In order to study this effect in isolation, it is necessary to construct model environments whereby all input variables except for $\Delta^{14}C$ are kept constant. We will make our aims clearer in the abstract and introduction of the final manuscript.

Indeed, there are many temporally dynamic variables which affect the distribution of single foraminifera ^{14}C values within pooled sediment core samples, including sedimentation rate, bioturbation depth, species abundance changes, local reservoir age, global $\Delta^{14}C$, foraminifera vulnerability to dissolution/breakage, number of foraminifera picked per sample, etc. The number of figures that we could have generated is, therefore, practically infinite, as the referee will appreciate. We do agree with the referee that it would be possible to add one or two more 'dynamic' scenarios for the benefit of the reader.

> I hence recommend that the authors run three different simulations to give it the attention it deserves and to further highlight the power of their approach: One with the suggested step-like function in sedimentation rate, one with the suggested change in species abundance, and one with both changes.

> This would help to understand what process impacts the multi-modal character of the true age population the most and in what way.

These two additional simulations would indeed add useful information for the reader when it comes to judging the effect of dynamic input variables and the creation of further age-depth artefacts. We did not want to make too many figures, hence we only included a dynamic simulation with a step in sedimentation rate. However, on second thought, it may be possible to include the helpful suggestion of the referee in the same figure by adding lines with different colours for each scenario.

> One key observation is that mean AMS 14C ages are generally younger than idealized mean 14C ages. The authors attribute this to the isotope mass balance effect. However, to me it seems that it must also be to some extent linked to the effect of bioturbation itself, essentially the character of the exponential pdf for true age (how extended and pointy it is), combined with character of the atmospheric calibration.

Whenever there is sample age heterogeneity of any type, the AMS ^{14}C age (the average of all $F^{14}C$ values of the foraminifera in the sample, which is subsequently converted to 14C yr) will always be younger than the idealised mean ^{14}C age. This effect is because the $F^{14}C$ scale (or pMC scale) is exponential vs time and is, therefore, an accurate reflection of the isotope mass balance of radioactive ^{14}C, whereas ^{14}C years are linear vs time. The offset between the AMS ^{14}C age and the idealised ^{14}C age increases as the radiocarbon heterogeneity contained within the sample increases – this can be attributed to e.g. reduced sedimentation rate, periods of dynamic $\Delta^{14}C$ etc, as the referee notes. We did discuss all these factors in depth in lines 200 to 225. We will try to improve the structure of that text in the final version.

> "Perfect" simulated sediment archive scenarios: I would argue that a "perfect" sediment core is one without and very little bioturbation (e.g., BD=1cm with high SR>10 cm). I think it is worth to rephrase to "most favorable" or "best-case".

Good point, will rephrase for clarity.

> Section 2.2. It needs to be specified what fraction of the total number of foraminifera has been picked by the simulations, maybe simply by inserting "with a fraction set by the operator".

In Section 2.2 we describe the methods in general, in 3.0 we describe the different runs, where the exact settings used (100% of forams picked or ignoring the 10% oldest forams) are stated. We will make this clearer in the final manuscript, as it is indeed jarring at the moment when reading the manuscript from beginning to end.

> I would expect that this fraction matters for the magnitude of offset between AMS mean 14C age and whole-sample/idealized 14C age. I think it is worth highlighting that this fraction in reality changes as the abundance of (well-preserved) foraminifera changes in sediment cores, and that this would affect the observations (in what way?).

Regarding the first sentence of the comment: this effect can be clearly seen in the results, when one compares Fig. 1a (0% broken forams) and Fig. 2a (10% broken forams) and sees that the offset is systematically less in the case of the latter. We may have failed to mentioned this in the text, and will check to make sure we do when preparing the final version of the manuscript. Regarding the second sentence of the comment: Yes, in practice the fraction of broken foraminifera can of course change through time in the field as a function of dissolution, foraminifera ecology, bioturabtion depth, sedimentation rate. This issue is alluded to but should indeed be made clearer.

> Line 155: it is unclear what is meant with “samples nearing the blank value”. I would argue that samples with a 14C age of 45 kyr BP have a much larger analytical uncertainty. Please clarify.

Yes, “samples nearing the blank value” is indeed vague text. We will include the following in the final manuscript for clarity: “ Specifically, when assigning measurement errors to synthetic AMS determinations, a ^{14}C determination of 1.0 $F^{14}C$ is assumed to have an error of 30 ^{14}C yr, and a determination with the $F^{14}C$ value $e^{(\text{blankvalue-1})/-8033}$ (i.e. one ^{14}C yr younger than the blank value) is assumed to have an error of 200 ^{14}C yr. Errors (in ^{14}C yr) for intermediate dates are linearly interpolated to $F^{14}C$. ”

Regarding the choice of error value: 200 ^{14}C yr can indeed be considered an optimistic (but not unrealistic) measurement error for large (>1 mg) carbonate samples of ~46000 ^{14}C yr BP measured as graphite targets on the latest AMS systems. However, in our simulations we assume “best-case” conditions, thus also best case conditions for AMS measurement.

The remaining, more minor comments by the referee will also be addressed in the manuscript.

We’d like to thank the referee again for their extensive and helpful review.